

APPENDIX A

This directory includes the source files used to build test programs supporting the patent application, "Method for fast and efficient record retrieval in large databases."

```

fastdb.mk - make file to build programs.
fastdb.h - header file for programs.
mkdata.C - program to generate test Data file
mkindx.C - program to create index file (InsertRecord)
getrec.C - main test program to retrieve records
query.C - query record routines (3 methods)
crc.C - table driven crc routines for CRC-CCITT and CRC-16

```

To build programs:

```
$ make -f fastdb.mk      (uses g++ to compile)
```

To build an Data file with 1,000,000 records:

```
$ mkdata 1000000 > Data
```

To build the Index file:

```
$ mkindx Data
```

To test 10,000 random queries of 1,000,000 records:

```
$ timex getrec -s -l 10000 -r 1000000
```

Or to query a specific record:

```
$ getrec 0001000
```

The "gethash" and "getdirect" commands are built. These work the same as "getrec", but implement different search methods that were used in the test results of the patent application.

The test programs make a few assumptions:

- 1) The Data file is named "Data" and fields are <tab> separated.
- 2) The index file is named "keyindx".
- 3) The first field of "Data" file is the indexed "key" field.
- 4) The second field of "Data" file is decimal version of the first field. This is used as a validation in getrec to verify that the correct record was retrieved.
- 5) The parsing of Data record is left to application.
- 6) The Data schema file (for named fields) is not supported.
- 7) The Index file was created on same machine as getrec is run, ie, no provisions are made for machine dependant byte order of integers.
- 8) The CRCBUCKETSIZE (number of entries in an index bucket) should be tunable and stored in the index file (or somewhere). Currently, it is a #define in fastdb.h
- 9) The index field is assumed to be unique, the query function finds the first match and returns, it does not look for multiple records that match.